

<sup>1</sup> Fitness tracking reveals task-specific associations  
<sup>2</sup> between memory, mental health, and physical activity

<sup>3</sup> Jeremy R. Manning<sup>1, \*</sup>, Gina M. Notaro<sup>1</sup>, Esme Chen<sup>1</sup>, and Paxton C. Fitzpatrick<sup>1</sup>

<sup>4</sup> <sup>1</sup>Dartmouth College, Hanover, NH

<sup>5</sup> \*Address correspondence to jeremy.r.manning@dartmouth.edu

<sup>6</sup> July 16, 2022

7                   **Abstract**

8         Physical activity can benefit both physical and mental well-being. Different forms of exercise  
9         (e.g., aerobic versus anaerobic; running versus walking, swimming, or yoga; high-intensity  
10       interval training versus endurance workouts; etc.) impact physical fitness in different ways. For  
11       example, running may substantially impact leg and heart strength but only moderately impact  
12       arm strength. We hypothesized that the mental benefits of physical activity might be similarly  
13       differentiated. We focused specifically on how different intensities of physical activity might  
14       relate to different aspects of memory and mental health. To test our hypothesis, we collected (in  
15       aggregate) roughly a century's worth of fitness data. We then asked participants to fill out surveys  
16       asking them to self-report on different aspects of their mental health. We also asked participants to  
17       engage in a battery of memory tasks that tested their short and long term episodic, semantic, and  
18       spatial memory performance. We found that participants with similar physical activity habits  
19       and fitness profiles tended to also exhibit similar mental health and task performance profiles.  
20       These effects were task-specific in that different physical activity patterns or fitness characteristics  
21       varied with different aspects of memory, on different tasks. Taken together, these findings provide  
22       foundational work for designing physical activity interventions that target specific components  
23       of cognitive performance and mental health by leveraging low-cost fitness tracking devices.

24                   **Introduction**

25       Engaging in physical activity (exercise) can improve physical fitness by increasing muscle strength  
26       [9, 20, 23, 34], bone density [1, 8, 21], cardiovascular performance [24, 32], lung capacity [22,  
27       although see [35]], and endurance [43]. Physical activity can also improve mental health [2, 4, 10,  
28       12, 14, 25, 26, 27, 31, 33, 40] and cognitive performance [2, 3, 6, 11].

29       The physical benefits of exercise can be explained by stress-responses of the affected body  
30       tissues. For example, skeletal muscles that are taxed during exercise exhibit stress responses [28]  
31       that can in turn affect their growth or atrophy [36]. By comparison, the benefits of physical  
32       activity on mental health are less direct. For example, one hypothesis is that physical activity  
33       leads to specific physiological changes, such as increased aminergic synaptic transmission and

34 endorphin release, which in turn act on neurotransmitters in the brain [31]. Speculatively, if  
35 different physical activity regimens lead to different neurophysiological responses, one might be  
36 able to map out a spectrum of signalling and transduction pathways that are impacted by a given  
37 type, duration, and intensity of physical activity in each brain region. For example, prior work has  
38 shown that physical activity increases acetylcholine levels, starting in the vicinity of the exercised  
39 muscles [37]. Acetylcholine is thought to play an important role in memory formation [e.g., by  
40 modulating specific synaptic inputs from entorhinal cortex to the hippocampus, albeit in rodents;  
41 30]. Given the central role that these medial temporal lobe structures play in memory, changes in  
42 acetylcholine might lead to specific changes in memory formation and retrieval.

43 In the present study, we hypothesize that (a) different intensities of physical activity will have  
44 different, quantifiable impacts on cognitive performance and mental health, and that (b) these  
45 impacts will be consistent across individuals. To this end, we collected a year of real-world fitness  
46 tracking data from each of 113 participants. We then asked each participant to fill out a brief survey  
47 in which they self-evaluated and self-reported several aspects of their mental health. Finally, we  
48 ran each participant through a battery of memory tasks, which we used to evaluate their memory  
49 performance along several dimensions. We searched the data for potential associations between  
50 memory, mental health, and physical activity.

## 51 **Methods**

52 We ran an online experiment using the Amazon Mechanical Turk (MTurk) platform [13]. We  
53 collected data about each participant’s fitness and physical activity habits, a variety of self-reported  
54 measures concerning their mental health, and about their performance on a battery of memory  
55 tasks.

56 **Experiment**

57 **Participants**

58 We recruited experimental participants by posting our experiment as a Human Intelligence Task  
59 (HIT) on the MTurk platform. We limited participation to MTurk Workers who had been assigned  
60 a “master worker” designation on the platform, given to workers who score highly across several  
61 metrics on a large number of HITs, according to a proprietary algorithm managed by Amazon.  
62 One criterion embedded into the algorithm is a requirement that master workers must maintain a  
63 HIT acceptance rate of at least 95%. We further limited our participant pool to participants who  
64 self-reported that they were fluent in English and regularly used a Fitbit fitness tracker device.  
65 A total of 160 workers accepted our HIT in order to participate in our experiment. Of these,  
66 we excluded all participants who failed to log into their Fitbit account (giving us access to their  
67 anonymized fitness tracking data), encountered technical issues (e.g., by accessing the HIT using an  
68 incompatible browser, device, or operating system), or who ended their participation prematurely,  
69 before completing the full study. In all, 113 participants contributed usable data to the study.

70 For their participation, workers received a base payment of \$5 per hour (computed in 15  
71 minute increments, rounded up to the nearest 15 minutes), plus an additional performance-based  
72 bonus of up to \$5. Our recruitment procedure and study protocol were approved by Dartmouth’s  
73 Committee for the Protection of Human Subjects. We obtained informed consent using an online  
74 form administered to all prospective participants prior to enrolling them in our study. All methods  
75 were performed in accordance with the relevant guidelines and regulations.

76 **Gender, age, and race.** Of the 113 participants who contributed usable data, 77 reported their  
77 gender as female, 35 as male, and 1 chose not to report their gender. Participants ranged in age  
78 from 19 to 68 years old (25<sup>th</sup> percentile: 28.25 years; 50<sup>th</sup> percentile: 32 years; 75<sup>th</sup> percentile: 38  
79 years). Participants reported their race as White (90 participants), Black or African American (11  
80 participants), Asian (7 participants), Other (4 participants), and American Indian or Alaska Native  
81 (3 participants). One participant opted not to report their race.

82 **Languages.** All participants reported that they were fluent in either 1 and 2 languages (25<sup>th</sup>  
83 percentile: 1; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 1), and that they were “familiar” with between 1  
84 and 11 languages (25<sup>th</sup> percentile: 1; 50<sup>th</sup> percentile: 2; 75<sup>th</sup> percentile: 3).

85 **Reported medical conditions and medications.** Participants reported having and/or taking med-  
86 ications pertaining to the following medical conditions: anxiety or depression (4 participants),  
87 recent head injury (2 participants), high blood pressure (1 participant), bipolar disorder (1 partici-  
88 pant), hypothyroidism (1 participant), and other unspecified conditions or medications (1 partici-  
89 pant). Participants reported their current and typical stress levels on a Likert scale as very relaxed  
90 (-2), a little relaxed (-1), neutral (0), a little stressed (1), or very stressed (2). The “current” stress  
91 level reflected participants’ stress at the time they participated in the experiment. Their responses  
92 ranged from -2 to 2 (current stress: 25<sup>th</sup> percentile: -2; 50<sup>th</sup> percentile: -1; 75<sup>th</sup> percentile: 1; typical  
93 stress: 25<sup>th</sup> percentile: 0; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 1). Participants also reported their  
94 current level of alertness on a Likert scale as very sluggish (-2), a little sluggish (-1), neutral (0), a  
95 little alert (1), or very alert (2). Their responses ranged from -2 to 2 (25<sup>th</sup> percentile: 0; 50<sup>th</sup> per-  
96 centile: 1; 75<sup>th</sup> percentile: 2). Nearly all (111 out of 113) participants reported that they had normal  
97 color vision, and 15 participants reported uncorrected visual impairments (including dyslexia and  
98 uncorrected near- or far-sightedness).

99 **Residence and level of education.** Participants reported their residence as being located in the  
100 suburbs (36 participants), a large city (30 participants), a small city (23 participants), rural (14 partic-  
101 ipants), or a small town (10 participants). Participants reported their level of education as follows:  
102 College graduate (42 participants), Master’s degree (23 participants), Some college (21 partici-  
103 pants), High school graduate (9 participants), Associate’s degree (8 participants), Other graduate  
104 or professional school (5 participants), Some graduate training (3 participants), or Doctorate (2  
105 participants).

106 **Reported water and coffee intake.** Participants reported the number of 8 oz cups of water and  
107 coffee they had consumed prior to accepting the HIT. Water consumption ranged from 0 to 6 cups

<sup>108</sup> (25<sup>th</sup> percentile: 1; 50<sup>th</sup> percentile: 3; 75<sup>th</sup> percentile: 4). Coffee consumption ranged from 0 to 4  
<sup>109</sup> cups (25<sup>th</sup> percentile: 0; 50<sup>th</sup> percentile: 1; 75<sup>th</sup> percentile: 2).

<sup>110</sup> **Tasks**

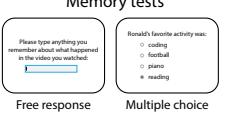
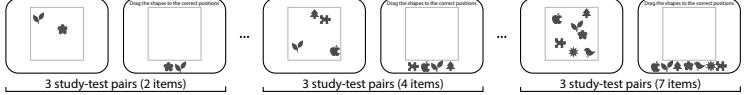
<sup>111</sup> Upon accepting the HIT posted on MTurk, each worker was directed to read and fill out a screening  
<sup>112</sup> and consent form, and to share access to their anonymized Fitbit data via their Fitbit account. After  
<sup>113</sup> consenting to participate in our study and successfully sharing their Fitbit data, participants filled  
<sup>114</sup> out a survey and then engaged in a series of memory tasks (Fig. 1). All stimuli and code for running  
<sup>115</sup> the full MTurk experiment may be found [here](#).

<sup>116</sup> **Survey questions.** We collected the following demographic information from each participant:  
<sup>117</sup> their birth year, gender, highest (academic) degree achieved, race, language fluency, and language  
<sup>118</sup> familiarity. We also collected information about participants' health and wellness, including about  
<sup>119</sup> their vision, alertness, stress, sleep, coffee and water consumption, location of their residence,  
<sup>120</sup> activity typically required for their job, and physical activity habits.

<sup>121</sup> **Free recall (Fig. 1a).** Participants studied a sequence of four word lists, each comprising 16 words.  
<sup>122</sup> After studying each list, participants received an immediate memory test, whereby they were asked  
<sup>123</sup> to type (one word at a time) any words they remembered from the just-studied list, in any order.

<sup>124</sup> Words were presented for 2 s each, in black text on a white background, followed by a 2 s blank  
<sup>125</sup> (white) screen. After the final 2 s pause, participants were given 90 s to type in as many words  
<sup>126</sup> as they could remember, in any order. The memory test was constructed such that the participant  
<sup>127</sup> could only see the text of the current word they were typing; when they pressed any non-letter  
<sup>128</sup> key, the current word was submitted and the text box they were typing in was cleared. This was  
<sup>129</sup> intended to prevent participants from retroactively editing their previous responses.

<sup>130</sup> The word lists participants studied were drawn from the categorized lists reported by [44]. Each  
<sup>131</sup> participant was assigned four unique randomly chosen lists (in a randomized order), selected from  
<sup>132</sup> a full set of 16 lists. Each chosen list was then randomly shuffled before presenting the words to

	Main task and immediate memory test				Delayed memory test
a.	1 Free recall	Study words 	Memory test 		5 
b.	2 Naturalistic recall	Watch a short video (The Temple of Knowledge) 	Memory tests 		6 
c.	3 Foreign language flashcards	Study flashcards 	Memory test 		7 
d.	4 Spatial learning	Memorize the positions of increasing numbers of shapes 			N/A

**Figure 1: Battery of memory tasks.** **a. Free recall.** Participants study 16 words (presented one at a time), followed by an immediate memory test where they type each word they remember from the just-studied list. In the delayed memory test, participants type any words they remember studying, from any list. **b. Naturalistic recall.** Participants watch a brief video, followed by two immediate memory tests. The first test asks participants to write out what happened in the video. The second test has participants answer a series of multiple choice questions about the conceptual content of the video. In the delayed memory test, participants (again) write out what happened in the video. **c. Foreign language flashcards.** Participants study a sequence of 10 English-Gaelic word pairs, each presented with an illustration of the given word. During an immediate memory test, participants perform a multiple choice test where they select the Gaelic word that corresponds to the given photograph. During the delayed memory test, participants perform a second multiple choice test, where they select the Gaelic word that corresponds to each of a new set of photographs. **d. Spatial learning.** In each trial, participants study a set of randomly positioned shapes. Next, the shapes' positions are altered, and participants are asked to drag the shapes back to their previous positions. **All panels.** The gray numbers denote the order in which participants experienced each task or test.

<sup>133</sup> the participants. Participants also performed a final delayed memory test where they were given  
<sup>134</sup> 180 s to type out any words they remembered from *any* of the 4 lists they had studied.

<sup>135</sup> Recalled words within an edit distance of 2 (i.e., a Levenshtein Distance less than or equal to  
<sup>136</sup> 2) of any word in the wordpool were “autocorrected” to their nearest match. We also manually  
<sup>137</sup> corrected clear typos or misspellings by hand (e.g., we corrected “hippoptumas” to “hippopota-  
<sup>138</sup> mus”, “zucinni” to “zucchini”, and so on). Finally, we lemmatized each submitted word to match  
<sup>139</sup> the plurality of the matching wordpool word (e.g., “bongo” was corrected to “bongos”, and so  
<sup>140</sup> on). After applying these corrections, any submitted words that matched words presented on the  
<sup>141</sup> just-studied list were tagged as “correct” recalls, and any non-matching words were discarded  
<sup>142</sup> as “errors.” Because participants were not allowed to edit the text they entered, we chose not to  
<sup>143</sup> analyze these putative “errors,” since we could not distinguish typos from true misrememberings.

<sup>144</sup> **Naturalistic recall (Fig. 1b).** Participants watched a 2.5-minute video clip entitled “The Temple  
<sup>145</sup> of Knowledge.” The video comprises an animated story told to StoryCorps by Ronald Clark, who  
<sup>146</sup> was interviewed by his daughter, Jamilah Clark. The narrator (Ronald) discusses growing up  
<sup>147</sup> living in an apartment over the Washington Heights branch of the New York Public Library, where  
<sup>148</sup> his father worked as a custodian during the 1940s.

<sup>149</sup> After watching the video clip, participants were asked to type out anything they remembered  
<sup>150</sup> about what happened in the video. They typed their responses into a text box, one sentence at a  
<sup>151</sup> time. When the participant pressed the return key or typed any final punctuation mark (“.”, “!”, or  
<sup>152</sup> “?”) the text currently entered into the box was “submitted” and added to their transcript, and the  
<sup>153</sup> text box was cleared to prevent further editing of any already-submitted text. This was intended to  
<sup>154</sup> prevent participants from retroactively editing their previous responses. Participants were given  
<sup>155</sup> up to 10 minutes to enter their responses. After 4 minutes, participants were given the option of  
<sup>156</sup> ending the response period early, e.g., if they felt they had finished entering all the information  
<sup>157</sup> they remembered. Each participant’s transcript was constructed from their submitted responses by  
<sup>158</sup> combining the sentences into a single document and removing extraneous whitespace characters.  
<sup>159</sup> Following this 4–10-minute free response period, participants were given a series of 10 multiple

<sup>160</sup> choice questions about the conceptual content of the story. All participants received the same  
<sup>161</sup> questions, in the same order. Participants also performed a final delayed memory test, where they  
<sup>162</sup> carried out the free response recall task a second time, near the end of the testing session. This  
<sup>163</sup> resulted in a second transcript, for each participant.

<sup>164</sup> **Foreign language flashcards (Fig. 1c).** Participants studied a series of 10 English-Gaelic word  
<sup>165</sup> pairs in a randomized order. We selected the Gaelic language both for its relatively small number  
<sup>166</sup> of native speakers and for its dissimilarity to other commonly spoken languages amongst MTurk  
<sup>167</sup> workers. We verified (via self report) that all of our participants were fluent in English and that  
<sup>168</sup> they were neither fluent nor familiar with Gaelic.

<sup>169</sup> Each word's "flashcard" comprised a cartoon depicting the given word, the English word or  
<sup>170</sup> phrase in lowercase text (e.g., "the boy"), and the Gaelic word or phrase in uppercase text (e.g.,  
<sup>171</sup> "BUACHAILL"). Each flashcard was displayed for 4 s, followed by a 3 s interval (during which  
<sup>172</sup> the screen was cleared) prior to the next flashcard presentation.

<sup>173</sup> After studying all 10 flashcards, participants were given a multiple choice memory test where  
<sup>174</sup> they were shown a series of novel photographs, each depicting one of the 10 words they had  
<sup>175</sup> learned. They were asked to select which (of 4 unique options) Gaelic word went with the given  
<sup>176</sup> picture. The 3 incorrect options were selected at random (with replacement across trials), and the  
<sup>177</sup> order in which the choices appeared to the participant were also randomized. Each of the 10 words  
<sup>178</sup> they had learned were tested exactly once.

<sup>179</sup> Participants also performed a final delayed memory test, where they were given a second set of  
<sup>180</sup> 10 questions (again, one per word they had studied). For this second set of questions participants  
<sup>181</sup> were prompted with a new set of novel photographs, and new randomly chosen incorrect choices  
<sup>182</sup> for each question. Each of the 10 original words they had learned were (again) tested exactly once  
<sup>183</sup> during this final memory test.

<sup>184</sup> **Spatial learning (Fig. 1d).** Participants performed a series of study-test trials where they memo-  
<sup>185</sup> rized the onscreen spatial locations of two or more shapes. During the study phase of each trial,

186 a set of shapes appeared on the screen for 10 s, followed by 2 s of blank (white) screen. During the  
187 test phase of each trial, the same shapes appeared onscreen again, but this time they were vertically  
188 aligned and sorted horizontally in a random order. Participants were instructed to drag (using the  
189 mouse) each shape to its studied position, and then to click a button to indicate that the placements  
190 were complete.

191 In different study-test trials, participants learned the locations of different numbers of shapes  
192 (always drawn from the same pool of 7 unique shapes, where each shape appeared at most one  
193 time per trial). They first performed three trials where they learned the locations of 2 shapes; next  
194 three trials where they learned the locations of 3 shapes; and so on until their last three trials, where  
195 (during each trial) they learned the locations of 7 shapes. All told, each participant performed 18  
196 study-test trials of this spatial learning task (3 trials for each of 2, 3, 4, 5, 6, and 7 shapes).

#### 197 **Fitness tracking using Fitbit devices**

198 To gain access to our study, participants provided us with access to all data associated with their  
199 Fitbit account from the year (365 calendar days) up to and including the day they accepted the HIT.  
200 We filtered out all identifiable information (e.g., participant names, GPS coordinates, etc.) prior to  
201 importing their data.

#### 202 **Collecting and processing Fitbit data**

203 The fitness tracking data associated with participants' Fitbit accounts varied in scope and duration  
204 according to which device the participant owned (Fig. S1), how often the participant wore (and/or  
205 synced) their tracking device, and how long they had owned their device. For example, while all  
206 participants' devices supported basic activity metrics such as daily step counts, only a subset of  
207 the devices with heart rate monitoring capabilities provided information about workout intensity,  
208 resting heart rate, and other related measures. Across all devices, we collected the following infor-  
209 mation: heart rate data, sleep tracking data, logged bodyweight measurements, logged nutrition  
210 measurements, Fitbit account and device settings, and activity metrics.

211 **Heart rate.** If available, we extracted all heart rate data collected by participants' Fitbit device(s)  
212 and associated with their Fitbit profile. Depending on the specific device model(s) and settings, this  
213 included second-by-second, minute-by-minute, daily summary, weekly summary, and/or monthly  
214 summary heart rate information. These summaries include information about participants' aver-  
215 age heart rates, and the amount of time they were estimated to have spent in different "heart rate  
216 zones" (rest, out-of-range, fat burn, cardio, or peak, as defined by their Fitbit profile), as well as an  
217 estimate of the number of estimated calories burned while in each heart rate zone.

218 **Sleep.** If available, we extracted all sleep data collected by participants' Fitbit device(s). Depend-  
219 ing on the specific device model(s) and settings, this included nightly estimates of the duration  
220 and quality of sleep, as well as the amount of time spent in each sleep stage (awake, REM, light, or  
221 deep).

222 **Weight.** If available, we extracted any weight-related information affiliated with participants'  
223 Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific device  
224 model(s) and settings, this included their weight, body mass index, and/or body fat percentage.

225 **Nutrition.** If available, we extracted any nutrition-related information affiliated with participants'  
226 Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific account  
227 settings and usage behaviors, this included a log of the specific foods they had eaten (and logged)  
228 over the past year, and the amount of water consumed (and logged) each day.

229 **Account and device settings.** We extracted any settings associated with participants' Fitbit ac-  
230 counts to determine (a) which device(s) and model(s) are associated with their Fitbit account, (b)  
231 time(s) when their device(s) were last synced, and (c) battery level(s).

232 **Activity metrics.** If available, we extracted any activity-related information affiliated with par-  
233 ticipants' Fitbit accounts within 1 year prior to enrolling in our study. Depending on their specific  
234 device model(s) and settings, this included: daily step counts; daily amount of time spent in each

235 activity level (sedentary, lightly active, fairly active, or very active, as defined by their account  
236 settings and preferences); daily number of floors climbed; daily elevation change; and daily total  
237 distance traveled.

238 **Comparing recent versus baseline measurements.**

239 We were interested in separating out potential associations between *absolute* fitness metrics and  
240 *relative* metrics. To this end, in addition to assessing potential raw (absolute) fitness metrics, we  
241 also defined a simple measure of recent changes in those metrics, relative to a baseline:

$$\Delta_{R,B}m = \frac{B \sum_{i=1}^R m(i)}{R \sum_{i=R+1}^{R+B} m(i)},$$

242 where  $m(i)$  is the value of metric  $m$  from  $i - 1$  days prior to testing (e.g.,  $m(1)$  represents the value  
243 of  $m$  on the day the participant accepted the HIT, and  $m(10)$  represents the value of  $m$  9 days prior  
244 to accepting the HIT). We set  $R = 7$  and  $B = 30$ . In other words, to estimate recent changes in any  
245 metric  $m$ , we divided the average value of  $m$  taken over the prior week by the average value of  $m$   
246 taken over the 30 days before that.

247 **Exploratory correlation analyses**

248 We used a bootstrap procedure to identify reliable correlations between different memory-related,  
249 fitness-related, and demographic-related variables. For each of  $N = 10,000$  iterations, we selected  
250 (with replacement) a sample of 113 participants to include. This yielded, for each iteration, a  
251 sampled “data matrix” with one row per sampled participant and one column for each measured  
252 variable. When participants were sampled multiple times in a given iteration, as was often the  
253 case, this matrix contained duplicate rows. Next, we computed the Pearson’s correlation between  
254 each pair of columns. This yielded, for each pair of columns, a distribution of  $N$  bootstrapped  
255 correlation coefficients. If 97.5% or fewer of the coefficients for a given pair of columns had the  
256 same sign, we excluded the pair from further analysis and considered the expected correlation  
257 between those columns to be undefined. If > 97.5% of the coefficients for a given pair of columns

had the same sign (corresponding to a bootstrap-estimated two-tailed  $p$  threshold of 0.05), we computed the expected correlation coefficient as:

$$\mathbb{E}_{i,j}[r] = \tanh\left(\frac{1}{N} \sum_{n=1}^N \tanh^{-1}(\text{corr}(m(i)_n, m(j)_n))\right),$$

where  $m(x)_n$  represents column  $x$  of the bootstrapped data matrix for iteration  $n$ ,  $\tanh$  is the hyperbolic tangent, and  $\tanh^{-1}$  is the inverse hyperbolic tangent. We estimated the corresponding  $p$ -values for these correlations as one minus the proportion of bootstrapped correlations with the same sign, multiplied by two.

#### Reverse correlation analyses

We sought to characterize potential associations between the *dynamics* of participants' fitness-related activities leading up to the time they participated in a memory task and their performance on the given task. For each fitness-related variable, we constructed a timeseries matrix whose rows corresponded to timepoints (sampled once per day) leading up to the day the participant accepted the HIT for our study, and whose columns corresponded to different participants. These matrices often contained missing entries, since different participants' Fitbit devices tracked fitness-related activities differently. For example, participants whose Fitbit devices lacked heart rate sensors would have missing entries for any heart rate-related variables. Or, if a given participant neglected to wear their fitness tracker on a particular day, the column corresponding to that participant would have missing entries for that day. To create stable estimates, we smoothed the timeseries of each fitness measure using a sliding window of 1 week. In other words, for each fitness measure, we replaced the "observed value" for each day with the average values of that measure (when available) over the 7-day interval ending on the given day.

In addition to this set of matrices storing timeseries data for each fitness-related variable, we also constructed a memory performance matrix,  $M$ , whose rows corresponded to different memory-related variables, and whose columns corresponded to different participants. For example, one row of the memory performance matrix reflected the average proportion of words (across lists)

282 that each participant remembered during the immediate free recall test, and so on.

283 Given a fitness timeseries matrix,  $F$ , we computed the weighted average and weighted standard  
284 error of the mean of each row of  $F$ , where the weights were given by a particular memory-related  
285 variable (row of  $M$ ). For example, if  $F$  contained participants' daily step counts, we could use  
286 any row of  $M$  to compute a weighted average across any participants who contributed step count  
287 data on each day. Choosing a row of  $M$  that corresponded to participants' performance on the  
288 naturalistic recall task would mean that participants who performed better on the naturalistic recall  
289 task would contribute more to the weighted average timeseries of daily step counts. Specifically,  
290 for each row,  $t$ , of  $F$ , we computed the weighted average (across the  $S$  participants) as:

$$\bar{f}(t) = \sum_{s=1}^S \hat{m}(s)F(t,s),$$

291 where  $\hat{m}$  denotes the normalized min-max scaling of  $m$  (the row of  $M$  corresponding to the chosen  
292 memory-related variable):

$$\hat{m} = \frac{m}{\sum_{s=1}^S \hat{m}(s)},$$

293 where

$$\hat{m} = (1 - \epsilon) \frac{m - \min(m)}{\max(m) - \min(m)} + \epsilon.$$

294 Here,  $\epsilon$  provides a lower bound on the influence of the lowest-weighted participant's data. We  
295 defined  $\epsilon = 0.001$ , ensuring that the lowest-weighted participant had relatively low (but non-zero)  
296 influence. We computed the weighted standard error of the mean as:

$$\text{SEM}_m(f(t)) = \frac{\left| \sum_{s=1}^S (F(t,s) - \bar{f}(t)) \right|}{\sqrt{S}}.$$

297 When a given row of  $F$  was missing data from one or more participants, those participants were  
298 excluded from the weighted average for the corresponding timepoint and the weights (across all  
299 remaining participants) were re-normalized to sum to 1. The above procedure yielded, for each  
300 memory variable, a timeseries of weighted average (and weighted standard error of the mean)

301 fitness tracking values leading up to the day of the experiment.

## 302 Results

303 Before testing our main hypotheses, we examined the behavioral data from each of four memory  
304 tasks: a random word list learning “free recall” task (Fig. 1); a naturalistic recall task whereby par-  
305 ticipants watched a short video and then recounted the narrative; a foreign language “flashcards”  
306 task; and a spatial learning task. Each of the first three tasks (free recall, naturalistic recall, and the  
307 flashcards task) included both an immediate (short term) memory test and a delayed (long term)  
308 memory test. The spatial learning task included only an immediate test. Participants in all four  
309 tasks exhibited general trends and tendencies that have been previously reported in prior work.  
310 We were also interested in characterizing the variability in task performance across participants.  
311 For example, if all participants exhibited near-identical behaviors or performance on a given task,  
312 we would be unable to identify how memory performance on that task varied with mental health  
313 or physical activity.

314 When participants engaged in free recall of random word lists, they displayed strong primacy  
315 and recency effects [29] on the immediate memory tests (as reflected by improved memory for  
316 early and late list items; Fig. 2a, left and right panels). On the delayed memory test, the recency  
317 effect was substantially diminished (Fig. 3a, left and right panels), consistent with myriad previous  
318 studies [for review see 18]. Participants also tended to cluster their recalls according to the words’  
319 study positions [17] on both the immediate (Fig. 2a, middle panel) and delayed (Fig. 3a, middle  
320 panel) memory tests.

321 When participants engaged in naturalistic recall by recounting the narrative of a short story  
322 video, they reliably and accurately remembered the major narrative events on both the immediate  
323 (Fig. 2b) and delayed (Fig. 3b) tests. This is consistent with prior work showing that memory for  
324 rich narratives is both detailed and accurate [7, 15].

325 Performance on the foreign language flashcards task (immediate: Fig. 2c; delayed: Fig. 3c)  
326 varied substantially across participants, and did not show any clear serial position effects. Partic-

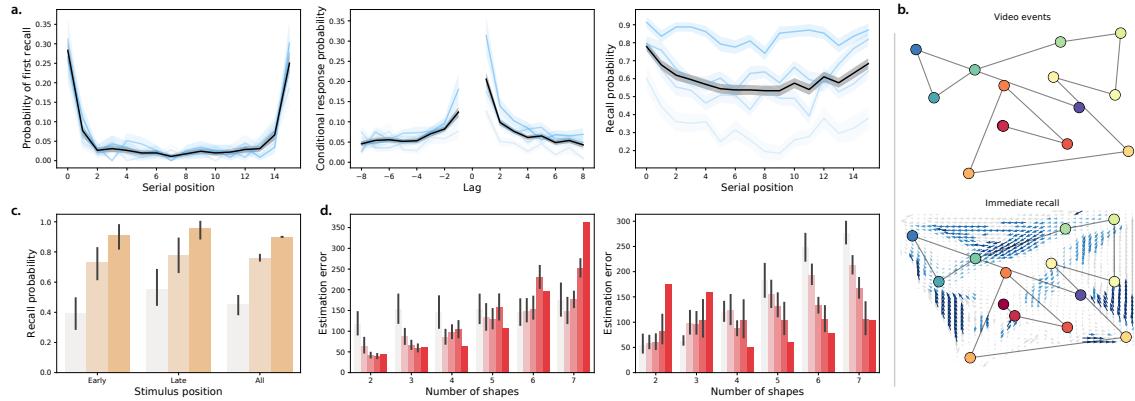
327 ipants also displayed substantial variation in performance on the spatial learning task (Fig. 2d).  
328 In general, participants reported the shape's positions more accurately when there were fewer  
329 shapes. However, both the baseline estimation accuracy and the rate of decrease in accuracy as a  
330 function of increasing number of memorized locations varied substantially across participants.

331 In addition to observing substantial across-participant variability in memory performance,  
332 we also observed substantial variability in participants' fitness and activity metrics (Fig. 4). We  
333 examined recent measurements, averaged over the week prior to testing (Fig. 4a), baselined mea-  
334 surements (average over the prior week, divided by the average over the preceding 30 days;  
335 Fig. 4b), along with more gradually varying measures that tended to remain relatively static over  
336 timescales of weeks to months (Fig. 4c). Figure S6 displays across-participant distributions for  
337 a broad selection of these measures, and Figures S7, S8, S9, and S10 show different participants'  
338 fitness metrics, broken down by their performance on different memory tasks.

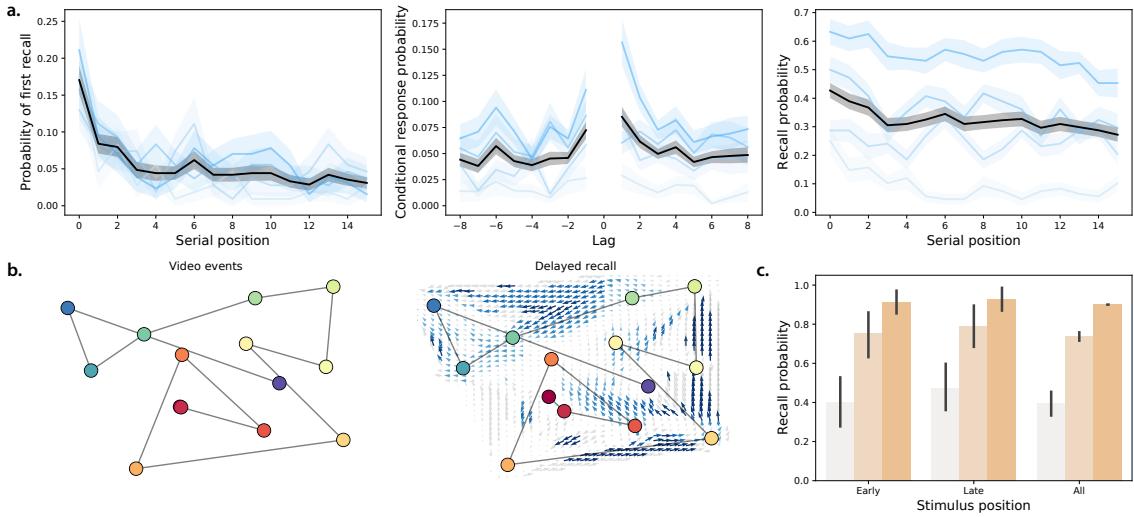
339 We wondered about potential links between the different aspects of participants' data. For  
340 example, if people who engaged in particular intensities of physical activity also tended to per-  
341 form better on a given memory task, this could suggest that either (a) some property intrinsic to  
342 participants who exercised in a particular way might also affect their memory performance on the  
343 given task, and/or (b) particular physical activity behaviors could have a causal impact on memory  
344 performance. We carried out an exploratory analysis whereby we used a bootstrap-based approach  
345 to identify reliable correlations between different aspects of memory performance (Fig. S11), dif-  
346 ferent aspects of fitness (Fig. S12), different demographic attributes (Fig. S13), and correlations  
347 between memory performance, fitness information, and demographic attributes (Fig. S14). Specif-  
348 ically, we sought to identify correlations that were present in the same direction (i.e., positive or  
349 negative) across different subsets of participants. For each test, we report the average correlation  
350 (taken across 10,000 subsets of participants, chosen with replacement) and an associated two-tailed  
351  $p$ -value, estimated as

$$p = 2 \times (1 - q),$$

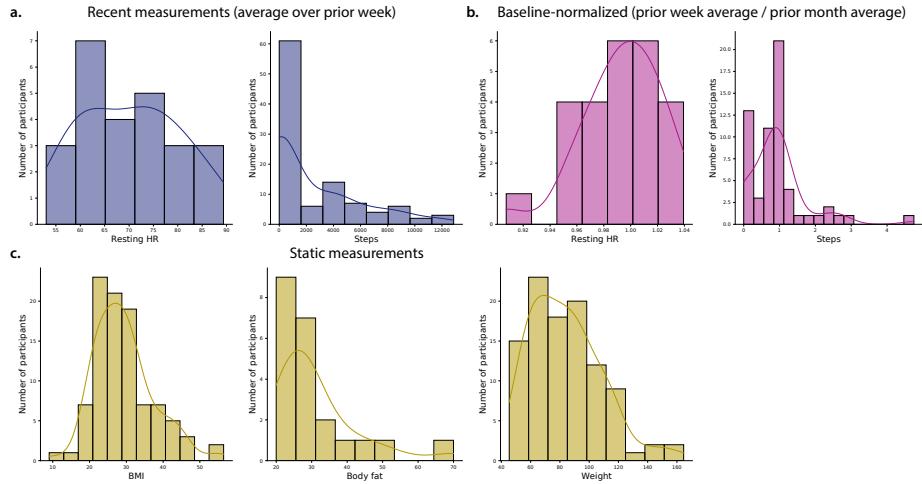
352 where  $q$  is the proportion of those 10,000 subsets that exhibited correlations in the same direction



**Figure 2: Immediate memory tests.** **a. Free recall.** Left: probability of recalling each word first as a function of its presentation position. Middle: probability of transitioning between successively recalling the word presented at position  $i$ , followed by word presented at position  $i + \text{Lag}$ . Right: probability of recalling each word as a function of its presentation position. See Figure S2 for additional details. **b. Naturalistic recall.** Top: 2D embedding of a 2.5-min video clip; each dot reflects a narrative event (red denotes early events and blue denotes later events). Bottom: 2D embedding of the averaged transcripts of participants' recounts (dots: same format as top panel). The arrows denote the average trajectory directions through the corresponding region of text embedding space, for any participants whose recounts passed through that region. Blue arrows denote statistically reliable agreement across participants ( $p < 0.05$ , corrected). See Figure S3 for additional details. **c. Foreign language flashcards.** Each bar denotes the average proportion of correctly recalled Gaelic-English word pairs from early (first 3), late (last 3), or all (i.e., all 10) study positions. See Figure S4 for additional details. **d. Spatial learning.** Average estimation error in shape locations as a function of the number of shapes. See Figure S5 for additional details. All panels: error bars and error ribbons denote bootstrap-estimated 95% confidence intervals. Shading (saturation) denotes results for different subsets of participants assigned based on their task performance (Figs. S2, S3, S4, and S5 provide information about which performance metrics and values the shading reflects; in general more saturated colors denote participants who performed better on the given task.) In Panel d, participants are grouped in two ways; in the left panel, participants are grouped according to the  $y$ -intercepts of regression lines (estimation error as a function of the number of shapes); in the right panel, participants are grouped according to the slopes of the same regression lines.



**Figure 3: Delayed memory tests.** **a. Free recall.** These panels are in the same format as Figure 2a, but they reflect performance on the delayed free recall task. For additional details see Figure S2. **b. Naturalistic recall.** These panels are in the same format as Figure 2b, but the right panel reflects performance on the delayed naturalistic recall task. For additional details see Figure S3. **c. Foreign language flashcards.** This panel is in the same format as Figure 2c, but it reflects performance on the delayed flashcards test. For additional details see Figure S4.



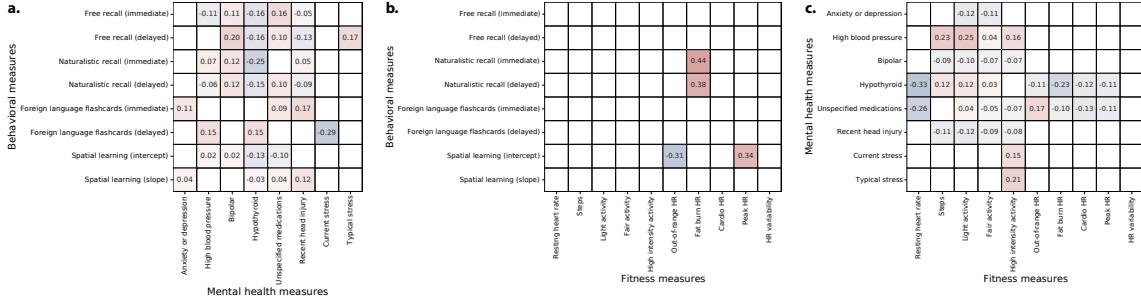
**Figure 4: Fitness measures.** **a. Recent measures.** Resting heart rate (HR) and daily step counts, averaged over the week prior to testing. **Baseline-normalized measures.** Resting heart rate and daily step counts averaged over the week prior to testing, divided by the average resting heart rate and step counts averaged over the preceding month. **Static measures.** Body mass index (BMI), body fat percentage, and weight (in kg). For more information see Figures S6, S7, S8, S9, and S10.

353 (see *Exploratory correlation analyses*). When all 10,000 randomly chosen subsets of participants ex-  
354 hibited correlations in the same direction (i.e., all positive correlations or all negative correlations),  
355 we report the  $p$ -value as  $p < 0.0001$ .

356 Several patterns emerged from these analyses. First, we found that participants' performance  
357 on the (within-task) immediate versus delayed memory tests from the free recall, naturalistic  
358 recall, and foreign language flashcards tasks were positively correlated ( $rs > 0.25, ps < 0.003$ ).  
359 This suggests that, within each of these tasks, similar processes or constraints may influence both  
360 short term and long term information retrieval. We also found reliable across-task correlations  
361 between participants' (immediate and delayed) performance on the free recall and foreign language  
362 flashcards tasks ( $rs > 0.3, ps < 0.03$ ).

363 A large number of fitness-related measures displayed reliable correlations (for a complete re-  
364 port, see Fig. S12). For example, body mass index (BMI) and weight were correlated ( $r = 0.91, p <$   
365  $0.0001$ ). Resting heart rate over the prior week was negatively correlated with recent low-to-  
366 moderate-intensity ("fat burn") cardiovascular activity levels ( $r = 0.70, p = 0.0004$ ). Participants'  
367 peak heart rates (averaged over the prior week) were also negatively correlated with recent in-  
368 creases in step counts and daily elevation gains ( $rs < -0.26, ps < 0.03$ ), where recent changes  
369 were defined as the average values over the seven days leading up to the test day divided by  
370 the average values over the preceding 30 days. Several demographic attributes (Fig. S13) dis-  
371 played trivial correlations (e.g., participants identifying as male never reported identifying as  
372 female, and so on). We also observed a negative correlation between reported stress and alertness  
373 ( $r = -0.44, p < 0.0001$ ), and positive correlations between the reported clarity of the instructions  
374 for all tasks ( $rs > 0.26, ps < 0.02$ ).

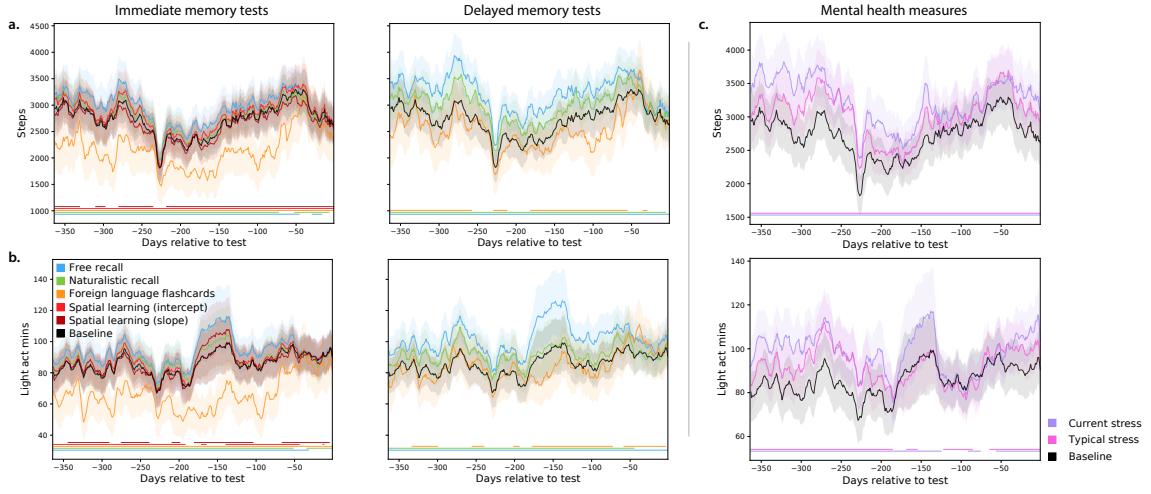
375 We also found reliable correlations between participants' fitness and demographic measures  
376 and their behaviors in different tasks (Fig. 5; for a complete report, see Fig. S14). For example,  
377 recent low-to-moderate-intensity ("fat burn") cardiovascular activity was positively correlated  
378 with immediate ( $r = 0.44, p = 0.001$ ) and delayed ( $r = 0.38, p = 0.031$ ) recall performance on the  
379 naturalistic memory task. Recent sedentary ("out-of-range") cardiovascular activity was negatively  
380 correlated with performance on the spatial learning task ( $r = -0.31, p = 0.042$ ), whereas recent



**Figure 5: Summaries of correlations between behavioral, fitness, and mental health measures.** The reported values in the tables reflect correlations between each behavioral measure and mental health measure. Only statistically reliable correlations ( $p < 0.05$ , corrected) are displayed. **a. Correlations between behavioral and mental health measures.** We adjusted each task's behavioral measure(s) such that more positive values reflect better performance on the given task. We used participants' mean recall accuracy to characterize performance on the free recall and foreign language flashcards tasks, and mean precision to characterize performance on the naturalistic recall tasks. We characterized performance on the spatial learning task using the (inverted and normalized) intercepts and slopes of linear regressions on mean estimation errors as a function of the number of studied shapes (also see Figs. 2, 3, S2, S3, S4, and S5). For each mental health measure, more positive values denote greater severity of the given measure. Typical and current stress levels were measured by self report. Mental health information was inferred using each participants' list of self-reported medications (see *Methods*). Positive correlations indicate that better performance on a given behavioral task is associated with more severe mental health phenotypes. **b. Correlations between fitness and mental health measures.** For each fitness measure, more positive values denote higher observed scores (i.e., higher resting heart rate, more minutes of activity or time spent in each heart rate zone, or greater heart rate variability). The mental health measures in this panel were treated as in Panel a. **c. Correlations between fitness and behavioral measures.** Each measure reflected in this panel was treated as in Panels a and b.

381 high intensity (“peak”) activity was positively correlated with performance on the spatial learning  
382 task ( $r = 0.34, p = 0.0002$ ). Mental health indicators, such as self-reported stress levels and  
383 medications were also associated with differences in memory (Figs. 5a, S14). For example, self-  
384 reported stress levels at the time of test were negatively correlated with performance on the delayed  
385 memory test for the foreign language flashcards task ( $r = -0.29, p = 0.038$ ), whereas participants  
386 who were medicated for anxiety and depression tended to perform slightly (but reliably) *better*  
387 on the immediate memory test for the foreign language flashcards task ( $r = 0.11, p < 0.0001$ ).  
388 Mental health indicators were also correlated with several fitness measures (Fig. 5c). For example,  
389 participants with higher resting heart rates were less likely to be hypothyroid ( $r = -0.33, p <$   
390  $0.0001$ ). Participants who engaged in more low-intensity (“light”) activity tended to be less anxious  
391 and depressed ( $r = -0.12, p = 0.03$ ), whereas participants who engaged in more high-intensity  
392 activity tended to report higher levels of current ( $r = 0.15, p = 0.027$ ) and typical ( $r = 0.21, p = 0.012$ )  
393 stress.

394 The above analyses indicate that recent differences in fitness-related activity are associated with  
395 differences in memory performance and mental health measures. Although the analyses treated  
396 these measures on average or in aggregate, many of the measures we collected are dynamic. For  
397 example, the amount or intensity of physical activity people engage in can vary over time, and  
398 so on. We wondered whether the dynamics of fitness-related measures might relate to memory  
399 performance and/or mental health measures. To this end, we carried out a series of reverse  
400 correlation analyses (see *Reverse correlation analyses*) to examine whether participants with different  
401 cognitive or mental health profiles also tended to display differences in fitness-related measures  
402 over time. In particular, we examined fitness data collected from participants’ Fitbit devices over the  
403 year prior to their test day in our study. Several example findings are summarized in Figure 6. We  
404 found that participants who performed well on the immediate and delayed free recall memory tests  
405 and on the naturalistic recall tests tended to be more active than participants who performed poorly  
406 on those tests (Figs. 6a, b; S15). Conversely, participants who performed well on the immediate  
407 and delayed foreign language flashcards tasks tended to be *less* active. These differences were  
408 present even a full year before the testing day. We also found substantial variability across people



**Figure 6: Dynamics of physical activity varies with memory performance and mental health measures.** **a. Daily step counts.** Each timecourse is weighted by either performance on immediate recall tests (left panel) or on delayed recall tests (right panel). The black (baseline) timecourses display the (unweighted) average across all participants. **b. Daily duration (in minutes) of low-intensity physical activity.** Timecourses are displayed in the same format and color scheme as those in Panel A. Analogous timecourses for additional fitness-related measures may be found in Figures S15, S16, and S17. **c. Timecourses of physical activity, weighted by mental health measures.** The timecourses in each panel display the average daily step counts (top panel) or duration of low-intensity activity (bottom panel). The colored lines show average activity dynamics weighted by self-reported stress levels at the start of the experiment (purple) and self-reported “typical” stress levels (pink). The baseline curves (black) display the average across all participants (re-plotted in Panel C to illustrate scale differences across panels). Timecourses for additional mental health-related and fitness-related measures may be found in Figures S18, S19, and S20. Error ribbons in all panels denote the standard error of the mean. Horizontal lines below each panel’s timecourses denote intervals over which each weighted measure (color) differs from the unweighted baseline (via a paired sample two-sided  $t$ -test of the weighted mean values for each measure within a 30-day window around each timepoint; horizontal lines denote  $p < 0.05$ , corrected).

409 with different (self-reported) mental health profiles (Figs. 6c, S18). Due to small sample sizes of  
410 individuals exhibiting several mental health dimensions, it is difficult to distinguish generalizable  
411 trends from individual differences that one or two individuals happened to exhibit. However,  
412 several large-sample-size trends emerged. For example, participants who reported higher levels  
413 of stress also tended to be slightly more physically active than participants who reported lower  
414 stress levels. We found analogous differences in other activity-related measures (Figs. S15 and S18),  
415 cardiovascular measures (Figs. S16 and S19), and sleep-related measures (Figs. S17 and S20). Taken  
416 together, the analyses suggest that cognitive and mental health differences are also associated with  
417 differences in the dynamics of physical health measures.

## 418 Discussion

419 After collecting a year's worth of fitness-tracking data from each of 113 participants, we ran each  
420 participant in a battery of memory tasks and had them fill out a series of demographic and mental  
421 health-related questions. We found that the associations between fitness-related activities, memory  
422 performance, and mental health are complex. For example, participants who tended to engage in  
423 a particular intensity of physical activity also tended to perform better on some memory tasks but  
424 worse on others. This suggests that engaging in one form or intensity of physical activity will not  
425 necessarily affect all aspects of cognitive or mental health equally (or in the same direction).

426 A number of prior studies have shown that engaging in exercise can improve cognitive and  
427 mental health [2, 3, 4, 6, 10, 11, 12, 14, 25, 26, 27, 31, 33, 40]. The majority of these studies ask  
428 participants in an “exercise intervention” condition (where participants engage in a designated  
429 physical activity or training regimen) or a “control” condition (where participants do not engage in  
430 the designated activity or training) to perform cognitive tasks or undergo mental health screening.  
431 In other words, most primary studies treat “physical activity” as a binary variable that either is  
432 or is not present for each participant. Most prior studies also track or manipulate exercise over  
433 relatively short durations (typically on the order of days or weeks). Our current work indicates that  
434 the true relations between physical activity, cognitive performance, and mental health may be non-

435 monotonic and heterogeneous across activities, tasks, and mental health measures. These relations  
436 can also unfold over much longer timescales than have been previously identified (on the order  
437 of months; Fig. 6). However, despite the complexities of the structures of these associations, we  
438 also found that they were often remarkably consistent across people. For example, as displayed  
439 in Figures 5 and S14, many of the associations between fitness, behavioral, and mental health  
440 measures were consistent across over 97.5% of 10,000 randomly chosen subsets of participants.

441 One important limitation of our study is that we cannot distinguish correlations between  
442 different measures from potential causal effects. For example, we cannot know (from our study)  
443 whether engaging in particular forms of physical activity *causes* changes in memory performance  
444 or mental health, or whether (alternatively) people who tend to engage in similar forms of physical  
445 activity also happen to exhibit similar memory and/or mental health profiles. In other words, an  
446 overlapping set of processes or person-specific attributes may lead someone to both form particular  
447 habits around physical activity and display high or low performance on a given memory test. We  
448 do not know whether memory performance or aspects of mental health might be manipulated  
449 or influenced by changing the patterns of physical activity someone engages in. For this reason,  
450 we have been careful to frame our findings as correlations and associations, rather than to imply  
451 knowledge about a causal direction of our findings.

452 Although the present study cannot reveal causal effects, a large prior literature provides some  
453 insight into potential causal effects by examining the neural and cognitive effects of a variety of  
454 exercise interventions [5, 16, 19, 38, 39, 41, 42]. A limitation of that prior work is that most of  
455 these studies examine how relatively short-term changes in physical activity (e.g., on timescales of  
456 hours to days or, rarely, weeks to months) affect a cognitive performance on single task or aspect  
457 of mental health. The present study examines longer-term physical activity (over a full year), and  
458 relates long-term physical activity history to performance on a variety of tasks and to a variety of  
459 mental health dimensions.

460 To the extent that physical activity *does* provide a non-invasive means of manipulating cog-  
461 nitive performance and mental health, our work may have exciting implications for cognitive  
462 enhancement. For example, one might imagine building a recommendation system that suggests

463 a particular physical activity regimen tailored to improve a specific aspect of an individual's cog-  
464 nitive performance (e.g., the efficacy of a student's study session for an upcoming exam) or mental  
465 health (e.g., reducing symptoms of anxiety before an important meeting). Just as strength training  
466 may be customized to target a specific muscle group, or to improve performance on a specific  
467 physical task, similar principles might also be applied to target specific improvements in cognitive  
468 fitness and mental health.

## 469 **Acknowledgements**

470 We acknowledge useful discussions with David Bucci, Emily Glasser, Andrew Heusser, Abigail  
471 Bartolome, Lorie Loeb, Lucy Owen, and Kirsten Ziman. Our work was supported in part by the  
472 Dartmouth Young Minds and Brains initiative, and by NSF grant number 2145172 to J.R.M. The  
473 content is solely the responsibility of the authors and does not necessarily represent the official  
474 views of our supporting organizations. This paper is dedicated to the memory of David Bucci, who  
475 helped to inspire the theoretical foundations of this work. Dave served as a mentor and colleague  
476 on the project prior to his passing.

## 477 **Data and code availability**

478 All analysis code and data used in the present manuscript may be found [here](#).

## 479 **Author contributions**

480 Concept: J.R.M. and G.M.N. Experiment implementation and data collection: G.M.N. Analyses:  
481 J.R.M., G.M.N., E.C., and P.C.F. Writing: J.R.M. with input from all authors.

## 482 **Competing interests**

483 The authors declare no competing interests.

484 **References**

- 485 [1] Bassey, E. J. and Ramsdale, S. J. (1994). Increase in femoral bone density in young women  
486 following high-impact exercise. *Osteoporosis International*, 4:72–75.
- 487 [2] Basso, J. C. and Suzuki, W. A. (2017). The effects of acute exercise on mood, cognition, neuro-  
488 physiology, and neurochemical pathways: a review. *Brain Plasticity*, 2(2):127–152.
- 489 [3] Brisswalter, J., Collardeau, M., and René, A. (2002). Effects of acute physical exercise charac-  
490 teristics on cognitive performance. *Sports Medicine*, 32:555–566.
- 491 [4] Callaghan, P. (2004). Exercise: a neglected intervention in mental health care? *Psychiatric and*  
492 *Mental Health Nursing*, 11(4):476–483.
- 493 [5] Chang, Y.-K., Chu, C.-H., Wang, C.-C., Wang, Y.-C., Song, T.-F., Tsai, C.-L., and Etnier, J. L.  
494 (2015). Dose-response relation between exercise duration and cognition. *Medicine and Science in*  
495 *Sports and Exercise*, 47:159–165.
- 496 [6] Chang, Y. K., Labban, J. D., Gapin, J. I., and Etnier, J. L. (2012). The effects of acute exercise on  
497 cognitive performance: a meta-analysis. *Brain Research*, 1453:87–101.
- 498 [7] Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2017). Shared  
499 memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*,  
500 20(1):115.
- 501 [8] Chilibeck, P. D., Sale, D. G., and Webber, C. E. (2012). Exercise and bone mineral density. *Sports*  
502 *Medicine*, 19:103–122.
- 503 [9] Crane, J. D., MacNeil, L. G., and Tarnopolsky, M. A. (2013). Long-term aerobic exercise is  
504 associated with greater muscle strength throughout the life span. *The Journals of Gerontology:*  
505 *Series A*, 68(6):631–638.
- 506 [10] Deslandes, A., Moraes, H., Ferreira, C., Veiga, H., Silveira, H., Mouta, R., Pompeu, F. A. M. S.,  
507 Coutinho, E. S. F., and Laks, J. (2009). Exercise and mental health: many reasons to move.  
508 *Neuropsychobiology*, 59:191–198.

- 509 [11] Etnier, J. L., Nowell, P. M., Landers, D. M., and Sibley, B. A. (2006). A meta-regression to  
510 examine the relationship between aerobic fitness and cognitive performance. *Brain Research: Brain Research Reviews*, 52(1):119–130.
- 511
- 512 [12] Gordon, B. R., McDowell, C. P., Lyons, M., and Herring, M. P. (2017). The effects of resistance exercise training on anxiety: a meta-analysis and meta-regression analysis of randomized controlled trials. *Sports Medicine*, 47(12):2521–2532.
- 513
- 514
- 515 [13] Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., and Chan, P. (2016). psiTurk: an open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, 48(3):829–842.
- 516
- 517
- 518 [14] Herring, M. P., O'Connor, P. J., and Dishman, R. K. (2010). The effects of exercise training on anxiety symptoms among patients: a systematic review. *Archives of Internal Medicine*, 170(4):321–331.
- 519
- 520
- 521 [15] Heusser, A. C., Fitzpatrick, P. C., and Manning, J. R. (2021). Geometric models reveal behavioral and neural signatures of transforming naturalistic experiences into episodic memories. *Nature Human Behavior*, 5:905–919.
- 522
- 523
- 524 [16] Imboden, C., Gerber, M., Beck, J., Eckert, A., Pühse, U., Holsboer-Trachsler, E., and Hatzinger, M. (2019). Effects of aerobic exercise as add-on treatment for inpatients with moderate to severe depression on depression severity, sleep, cognition, psychological well-being, and biomarkers: study protocol, description of study population, and manipulation check. *Frontiers in Psychiatry*, 10(262):doi.org/10.3389/fpsyg.2019.00262.
- 525
- 526
- 527
- 528
- 529 [17] Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition*, 24:103–109.
- 530
- 531 [18] Kahana, M. J. (2012). *Foundations of human memory*. Oxford University Press, New York, NY.
- 532 [19] Kamijo, K., Nishihira, Y., Higashiura, T., and Kuroiwa, K. (2007). The interactive effect of

- 533 exercise intensity and task difficulty on human cognitive processing. *International Journal of*  
534 *Psychophysiology*, 65(2):114–121.
- 535 [20] Knuttgen, H. G. (2007). Strength training and aerobic exercise: comparison and contrast.  
536 *Journal of Strength and Conditioning Research*, 21(3):973–978.
- 537 [21] Layne, J. E. and Nelson, M. E. (1999). The effects of progressive resistance training on bone  
538 density: a review. *Medicine and Science in Sports and Exercise*, 31(1):25–30.
- 539 [22] Lazovic-Popovic, B., Zlatkovic-Svenda, M., Durmic, T., Djelic, M., Saranovic, D., and Zugic, V.  
540 (2016). Superior lung capacity in swimmers: some questions, more answers! *Revista Portuguesa*  
541 *de Pneumologia*, 22(3):151–156.
- 542 [23] Lindh, M. (1979). Increase of muscle strength from isometric quadriceps exercises at different  
543 knee angles. *Scandinavian Journal of Rehabilitation Medicine*, 11(1):33–36.
- 544 [24] Maiorana, A., O'Driscoll, G., Cheetham, C., Collis, J., Goodman, C., Rankin, S., Taylor, R.,  
545 and Green, D. (2000). Combined aerobic and resistance exercise training improves functional  
546 capacity and strength in CHF. *Journal of Applied Physiology*, 88(1565–1570).
- 547 [25] Mikkelsen, K., Stojanovska, L., Polenakovic, M., Bosevski, M., and Apostolopoulos, V. (2017).  
548 Exercise and mental health. *Maturitas*, 106:48–56.
- 549 [26] Morres, I. D., Hatzigeorgiadis, A., Stathi, A., Comoutos, N., Arpin-Cribbie, C., Krommidas, C.,  
550 and Theodorakis, Y. (2018). Aerobic exercise for adult patients with major depressive disorder in  
551 mental health services: a systematic review and meta-analysis. *Depression and Anxiety*, 36(1):39–  
552 53.
- 553 [27] Morres, I. D., Tzouma, N.-A., Hatzigeorgiadis, A., Krommidas, C., Kotronis, K. V., Dafopoulos,  
554 K., Theodorakis, Y., and Comoutos, N. (2022). Exercise for perinatal depressive symptoms: a  
555 systematic review and meta-analysis of randomized controlled trials in perinatal health services.  
556 *Journal of Affective Disorders*, 298:26–42.

- 557 [28] Morton, J. P., Kayani, A. C., McArdle, A., and Drust, B. (2009). The exercise-induced stress  
558 response of skeletal muscle, with specific emphasis on humans. *Sports Medicine*, 39:643–662.
- 559 [29] Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology:*  
560 *General*, 64:482–488.
- 561 [30] Palacios-Filardo, J., Udakis, M., Brown, G. A., Tehan, B. G., Congreve, M. S., Nathan, P. J.,  
562 Brown, A. J. H., and Mellor, J. R. (2021). Acetylcholine prioritises direct synaptic inputs from  
563 entorhinal cortex to CA1 by differential modulation of feedforward inhibitory circuits. *Nature  
Communications*, 12(5475):doi.org/10.1038/s41467-021-25280-5.
- 565 [31] Paluska, S. A. and Schwenk, T. L. (2000). Physical activity and mental health. *Sports Medicine*,  
566 29(3):167–180.
- 567 [32] Pollock, M. L., Franklin, B. A., Balady, G. J., Chaltman, B. L., Fleg, J. L., Fletcher, B., Limacher,  
568 M., na, I. L. P., Stein, R. A., Williams, M., and Bazzarre, T. (2000). Resistance exercise in individuals  
569 with and without cardiovascular disease. *Circulation*, 101:828–833.
- 570 [33] Raglin, J. S. (1990). Exercise and mental health. *Sports Medicine*, 9:323–329.
- 571 [34] Rogers, M. A. and Evans, W. J. (1993). Changes in skeletal muscle with aging: effects of  
572 exercise training. *Exercise and Sport Sciences Reviews*, 21:65–102.
- 573 [35] Roman, M. A., Rossiter, H. B., and Casaburi, R. (2016). Exercise, ageing and the lung. *European  
574 Respiratory Journal*, 48:1471–1486.
- 575 [36] Schiaffino, S., Dyar, K. A., Ciciliot, S., Blaauw, B., and Sandri, M. (2013). Mechanisms regulating  
576 skeletal muscle growth and atrophy. *The febs Journal*, 280(17):4294–4314.
- 577 [37] Shoemaker, J. K., Halliwill, J. R., Hughson, R. L., and Joyner, M. J. (1997). Contributions of  
578 acetylcholine and nitric oxide to forearm blood flow at exercise onset and recovery. *Vascular  
579 Physiology*, 273(5):2388–2395.

- 580 [38] Sinha, N., Berg, C. N., Yassa, M. A., and Gluck, M. A. (2021). Increased dynamic flexibility in  
581 the medial temporal lobe network following an exercise intervention mediates generalization  
582 of prior learning. *Neurobiology of Learning and Memory*, 177:107340.
- 583 [39] Suwabe, K., Hyodo, K., Byun, K., Ochi, G., Yassa, M. A., and Soya, H. (2017). Acute moderate  
584 exercise improves mnemonic discrimination in young adults. *Hippocampus*, 27(3):229–234.
- 585 [40] Taylor, C. B., Sallis, J. F., and Needle, R. (1985). The relation of physical activity and exercise  
586 to mental health. *Public Health Reports*, 100(2):195–202.
- 587 [41] Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychologica*,  
588 112(3):297–324.
- 589 [42] Vidoni, E. D., Johnson, D. K., Morris, J. K., Van Sciver, A., Greer, C. S., Billinger, S. A.,  
590 Donnelly, J. E., and Burns, J. M. (2015). Dose-response of aerobic exercise on cognition: A  
591 community-based, pilot randomized controlled trial. *PLoS One*, 10(7):1–13.
- 592 [43] Wilmore, J. H. and Knuttgen, H. G. (2003). Aerobic exercise and endurance. *The Physician and*  
593 *Sportsmedicine*, 31(5):45–51.
- 594 [44] Ziman, K., Heusser, A. C., Fitzpatrick, P. C., Field, C. E., and Manning, J. R. (2018). Is  
595 automatic speech-to-text transcription ready for use in psychological experiments? *Behavior*  
596 *Research Methods*, 50:2597–2605.